Stochastic Processes and their Applications 3 (1975) 137–151

INFORMATION FLOW IN GRAPHS*

D.A. DAWSON

Carleton University, Ottawa, Canada

Received 8 August 1974
Revised 4 November 1974

A formula is obtained for the global flow of information in a discrete Markov system which is defined on a locally finite graph. It is shown that the information flow is related to a random walk on the lattice of finite subgraphs of the graph.

A.M.S. Subj. Class.: Primary 60K35; secondary 60J10.	
discrete Markov system	ergodic theory
information flow	graph

1. Introduction

A discrete Markov system is a stochastic model of a large system of interacting subsystems in which each elementary subsystem is a finite stochastic automata. There are two important classes of discrete Markov systems; synchronous systems and asynchronous systems. A synchronous system corresponds to the case in which all automata are permitted to undergo state transitions simultaneously. This is in contrast to the continuous time analogue in which case simultaneous state transitions of any two automata are forbidden. Although synchronous systems have quite naturally been used to model a number of biological phenomena (see Vasserstein [8]), it is demonstrated in [6] that synchronous systems cannot model many systems in thermodynamic equilibrium such as the two dimensional Ising model. In order to obtain a discrete time evolution consistent with the Ising-type model we introduced in [6] the concept of an asynchronous Markov system which may be thought of as the discrete time analogue of a continuous time system.

The main objective of the study of Markov systems is the determina-

* Research partially supported by the National Research Council of Canada and the Japan Society for the Promotion of Science.

tion of the global or macroscopic behavior of complex large systems. For example, if the internal noise is sufficiently intense then any structured organization inherent in the system will be dissipated and a single degenerate macroscopic behavior will be the result; such a system is known as an ergodic system. One approach to the study of this aspect of large systems theory is to analyze the flow of information throughout the system. In [3] a mathematical formulation of information flow in a discrete Markov system was first introduced and it was demonstrated that the ergodicity of such a system is a result of the absence of the flow of information over macroscopic distances. In [4] a more detailed study of the information flow in one dimensional systems was described and in [5] the information flow in two special classes of Markov systems was considered. In this paper we obtain a formula for the global flow of information in systems having an arbitrary locally finite interaction graph for both synchronous and asynchronous Markov systems. In addition we obtain a relationship between the information flow and an auxiliary Markov chain whose state space is the collection of finite subgraphs of the interaction graph and we show that the system is ergodic if the auxiliary Markov chain is subcritical.

Alternate approaches to the ergodic theory of discrete Markov systems have been developed by Chover [2] and Vasserstein [8].

2. Review of the basic concepts

Let A be a countable set, S a finite set and $\Gamma = S^A$. We can think of A as the set of labels of the elementary subsystems and we can think of Γ as the set of all configurations over A. For $\gamma \in \Gamma$, $\alpha \in A$, $Y_\alpha(\gamma) \equiv \gamma(\alpha)$ denotes the state of the subsystem labeled by α. For $B \subset A$, let $\mathcal{G}_B \equiv \sigma\{Y_\alpha : \alpha \in B\}$ and $\mathcal{G} \equiv \mathcal{G}_A$.

We next consider the construction of a Markov chain with state space Γ which will describe the time evolution of the system. Let $(\Omega, \mathcal{F}) \equiv (\{0, 1\}, \mathcal{F}_0) \otimes (\Gamma, \mathcal{G})^{\mathbf{Z}_+}$ where $\mathcal{F}_0$ is the Boolean algebra of subsets of $\{0, 1\}$ and $\mathbf{Z}_+ \equiv \{1, 2, 3, ...\}$. Given $\omega \in \Omega$, let $X_0(\omega) \equiv \omega(0) \in \{0, 1\}$ denote the index of the initial distribution. Given $\omega \in \Omega$, $n \geq 1$, let $X_{n,\alpha}(\omega) \equiv Y_\alpha(\omega(n))$, that is, the state of the αth subsystem at time n. For $J \subset A$, let

$$\mathcal{F}_J^n \equiv \sigma\{X_{n,\alpha} : \alpha \in J\} .$$

Given two probability measures m_0 and m_1 on $(\Gamma, \mathcal{G})$ and a probability transition kernel $P(\cdot,\cdot)$ on $(\Gamma, \mathcal{G})$ we can construct a measure $\mathbf{P}$ on $(\Omega, \mathcal{F})$ such that $\{X_n\}$ is a Markov Chain with

$$\mathbf{P}[X_0 = 0] = \mathbf{P}[X_0 = 1] = \tfrac{1}{2}\,, \tag{2.1}$$

$$\mathbf{P}[X_1 \in G \mid X_0 = 0] = m_0(G)\,, \tag{2.2}$$

$$\mathbf{P}[X_1 \in G \mid X_0 = 1] = m_1(G)\,, \tag{2.3}$$

and such that $X_1, X_2, X_3, \ldots$ is a time-homogeneous Markov chain with transition probabilities given by the kernel $\mathbf{P}[\cdot,\cdot]$.

We now restrict our attention to two important subclasses of such Markov chains for which there is an additional structure on A. We assume that A is a locally finite graph such that for each $\alpha \in A$, $N(\alpha)$ represents the finite subset of sites in A which are connected to α by an edge. In terms of systems we can think of $N(\alpha)$ as the set of inputs to the α^{th} subsystem and we call A the interaction graph. For each $\alpha \in A$, a mapping $q_\alpha : S^{N(\alpha)} \otimes S \to [0, 1]$ is a *local probability transition kernel* if for each $\zeta \in S^{N(\alpha)}$, $q_\alpha(\zeta, \cdot)$ is a probability mass function on S. The probability transition kernel $P(\cdot, \cdot)$ is said to be the *synchronous kernel* associated with the family $\{q_\alpha(\cdot, \cdot); \alpha \in A\}$ if for $y \in \Gamma$,

$$P[y, \cdot\,] = \prod_{\alpha \in A} q_\alpha(\Pi_{N(\alpha)}\, y, \cdot) \tag{2.4}$$

where $\Pi_C\, y$ denotes the restriction of the function y to the domain C.

Before describing the class of asynchronous systems we must introduce the notion of a totally disconnected random subset. Let $\mathcal{P}(A)$ denote the collection of all subsets of A and let $\mathcal{S}$ denote the σ-field of subsets of $\mathcal{P}(A)$ generated by events of the form

$$U(B) \equiv \{C\colon C \in \mathcal{P}(A),\ C \supset B\}\,,$$

$$L(B) \equiv \{C\colon C \in \mathcal{P}(A),\ C \cap B = \emptyset\}\,.$$

A *random subset* Ξ of A is a $\mathcal{P}(A)$-valued random variable. The distribution of a random subset is a probability measure Q defined on $(\mathcal{P}(A), \mathcal{S})$. The random subset is said to be *totally disconnected* if for any two finite sets D_1, D_2 with $D_1 \cap D_2 = \emptyset$,

$$Q(U(D_1) \cap L(D_2)) > 0$$

if and only if D_1 contains no pair α, β with $\alpha \in N(\beta)$.

We now define an asynchronous system as one in which simultaneous transitions of interconnected subsystems are forbidden and in which transitions occur on a totally disconnected random subset. Let $\{q_\alpha(\cdot): \alpha \in A\}$ be a family of local transition kernels and let $Q(\cdot)$ be the distribution of a totally disconnected random subset of $A \cdot Q(\cdot)$ will serve as the "speed measure" for the asynchronous system. The *asynchronous kernel* associated with $\{q_\alpha: \alpha \in A\}$ and $Q(\cdot)$ is defined by

$$\mathbf{P}[y, \{\gamma: Y_\alpha(\gamma) = x_\alpha, \alpha \in B\}] =$$

$$= \sum_{D \subset B} [Q(U(D) \cap L(B-D)) \prod_{\alpha \in D} q_\alpha(\Pi_{N(\alpha)} y, x_\alpha) \cdot \prod_{\alpha \in B-D} \delta(y_\alpha, x_\alpha) \tag{2.5}$$

for each finite set B and choice of $\{x_\alpha: \alpha \in B\}$.

Let $(\Omega, \mathcal{F}, \{X_n: n \geqslant 0\}, \mathbf{P})$ represent a discrete Markov system with local transition kernels $\{q_\alpha(\cdot,\cdot): \alpha \in A\}$. Consider finite subalgebras $\mathcal{F}_1$, $\mathcal{F}_2$, $\mathcal{F}_3$ and $\mathcal{F}_4$ of $\mathcal{F}$ and let $\mathcal{A}(\mathcal{F}_i)$ denote the set of atoms of $\mathcal{F}_i$. The *entropy* of $\mathcal{F}_i$ is defined by

$$I(\mathcal{F}_i) \equiv -\sum_{A \in \mathcal{A}(\mathcal{F}_i)} \mathbf{P}[A] \log(\mathbf{P}[A]), \tag{2.6}$$

where all logarithms are taken to base two. The *conditional entropy* in $\mathcal{F}_1$ given $\mathcal{F}_2$ is defined by

$$I(\mathcal{F}_1 \mid \mathcal{F}_2) \equiv -\sum_{B \in \mathcal{A}(\mathcal{F}_2)} \sum_{A \in \mathcal{A}(\mathcal{F}_1)} \mathbf{P}[A \cap B] \log(\mathbf{P}[A \mid B]). \tag{2.7}$$

The *mutual information* in the algebras $\mathcal{F}_1$ and $\mathcal{F}_2$ is defined by

$$I(\mathcal{F}_1 \wedge \mathcal{F}_2) \equiv I(\mathcal{F}_1) - I(\mathcal{F}_1 \mid \mathcal{F}_2), \tag{2.8}$$

and the mutual information in $\mathcal{F}_1$, $\mathcal{F}_2$ and $\mathcal{F}_3$ is defined by

$$I(\mathcal{F}_1 \wedge \mathcal{F}_2 \wedge \mathcal{F}_3) \equiv I(\mathcal{F}_1 \wedge \mathcal{F}_2) - I(\mathcal{F}_1 \wedge \mathcal{F}_2 \mid \mathcal{F}_3). \tag{2.9}$$

We also define *conditional mutual information* as follows:

$$I(\mathcal{F}_1 \wedge \mathcal{F}_2 \mid \mathcal{F}_3) \equiv I(\mathcal{F}_1 \mid \mathcal{F}_3) - I(\mathcal{F}_1 \mid \mathcal{F}_2 \vee \mathcal{F}_3), \tag{2.10}$$

$$I(\mathcal{F}_1 \wedge \mathcal{F}_2 \wedge \mathcal{F}_3 \mid \mathcal{F}_4) \equiv I(\mathcal{F}_1 \wedge \mathcal{F}_3 \mid \mathcal{F}_4) - I(\mathcal{F}_1 \wedge \mathcal{F}_2 \mid \mathcal{F}_3 \vee \mathcal{F}_4), \tag{2.11}$$

where $\mathcal{F}_2 \vee \mathcal{F}_3$ denotes the smallest algebra containing both $\mathcal{F}_2$ and $\mathcal{F}_3$. It is important to note that $I(\mathcal{F}_1 \wedge \mathcal{F}_2) \geqslant 0$, $I(\mathcal{F}_1 \wedge \mathcal{F}_2 \mid \mathcal{F}_3) \geqslant 0$ but that

$I(\mathcal{F}_1 \wedge \mathcal{F}_2 \wedge \mathcal{F}_3)$ and $I(\mathcal{F}_1 \wedge \mathcal{F}_2 \wedge \mathcal{F}_3 | \mathcal{F}_4)$ can be either positive or negative.

Let us again consider the discrete Markov system $(\Omega, \mathcal{F}, \{X_n: n \geq 0\}, P)$ and let the algebras $\mathcal{F}_0$ and $\mathcal{F}_J^n$ for $J \subset A$, $n \geq 1$, be defined as above. We now consider $\mathcal{F}^1_{N(\alpha)}$ as the input and $\mathcal{F}^2_{\{\alpha\}}$ as the output of a communications channel. The *transmission coefficient* of this channel is defined to be

$$\rho(\alpha) \equiv \operatorname{Sup} \frac{I(\mathcal{F}_0 \wedge \mathcal{F}^2_{\{\alpha\}})}{I(\mathcal{F}_0 \wedge \mathcal{F}^1_{N(\alpha)})} \tag{2.12}$$

where the supremum is taken over all choices of the initial measures m_0, m_1.

3. Global flow of information for synchronous systems

If a system has a number of different stable macroscopic behaviours, then these behaviours are determined by the initial conditions. The dependence of the subsystem corresponding to $J \subset A$ on the initial conditions is described by $I(\mathcal{F}_0 \wedge \mathcal{F}_J^n)$ which denotes the proportion of the "initial condition" information which is contained in the subsystem indexed by J at time n. It has been demonstrated in [4] that if $I(\mathcal{F}_0 \wedge \mathcal{F}_J^n) \to 0$ as $n \to \infty$ for all finite $J \subset A$, then the system is ergodic. In this section we derive an upper bound for $I(\mathcal{F}_0 \wedge \mathcal{F}_J^n)$.

Lemma 3.1. *Let $B_1 = \{\alpha_1, \alpha_2, ..., \alpha_r\}$ and $B_2 = \{\beta_1, \beta_2, ..., \beta_k\}$ be finite (possibly empty) subsets of A, $\alpha_0 \in A$, $\alpha_0 \notin B$ and $n \geq 1$. Then*

$$I(\mathcal{F}_0 \wedge \mathcal{F}^{n+1}_{\{\alpha_0\}} | \mathcal{F}^{n+1}_{B_1} \vee \mathcal{F}^{n+1}_{B_2}) \leq \rho\, I(\mathcal{F}_0 \wedge \mathcal{F}^n_{N(\alpha_0)} | \mathcal{F}^{n+1}_{B_1} \vee \mathcal{F}^n_{B_2}), \tag{3.1}$$

where we adopt the convention that

$$I(\mathcal{F}_0 \wedge \mathcal{F}^m_B | \mathcal{F}^{n+1}_{B_1} \vee \mathcal{F}^n_{B_2}) = \begin{cases} I(\mathcal{F}_0 \wedge \mathcal{F}^m_B | \mathcal{F}^{n+1}_{B_1}) & \text{if } B_2 = \emptyset \\ I(\mathcal{F}_0 \wedge \mathcal{F}^m_B) & \text{if } B_1 = B_2 = \emptyset. \end{cases}$$

Proof. Given $x_{\alpha_0} \in S$, $x_\alpha \in S$, $\alpha \in B_1$, $y_\beta \in S$, $\beta \in B_2 \cup N(\alpha_0)$ and $j \in \{0, 1\}$ it can be verified by a straightforward calculation that

$$\begin{aligned} &\mathrm{P}[X_{n+1,\alpha_0} = x_{\alpha_0} | X_0 = j; X_{n+1,\alpha} = x_\alpha, \alpha \in B_1; X_{n,\beta} = y_\beta; \beta \in B_2 \cup N(\alpha_0)] \\ &= q_{\alpha_0}(y_\beta, \beta \in N(\alpha_0); x_{\alpha_0}). \end{aligned} \tag{3.2}$$

This implies that, conditional on $\mathcal{F}^n_{N(\alpha_0)}$, $\mathcal{F}^{n+1}_{\{\alpha_0\}}$ is independent of

$$\mathcal{F}^{n+1}_{B_1} \vee \mathcal{F}^n_{B_2-N(\alpha_0)} \vee \mathcal{F}_0 \,.$$

It suffices to show that

$$I(\mathcal{F}^{n+1}_{\{\alpha_0\}} \wedge \mathcal{F}_0 \mid X_{n+1,\alpha} = x_\alpha, \alpha \in B_1; X_{n,\beta} = y_\beta, \beta \in B_2) \leqslant \tag{3.3}$$
$$\leqslant \rho(\alpha_0) I(\mathcal{F}^n_{N(\alpha_0)} \wedge \mathcal{F}_0 \mid X_{n+1} = x_\alpha, \alpha \in B_1; X_{n,\beta} = y_\beta, \beta \in B_2)\,.$$

But (3.2) demonstrates that even conditional on the event

$$\{X_{n+1,\alpha} = x_\alpha, \alpha \in B_1: X_{n,\beta} = y_\beta, \beta \in B_2\}$$

the channel from $\mathcal{F}^n_{N(\alpha_0)}$ to $\mathcal{F}^{n+1}_{\{\alpha_0\}}$ is defined by $q_{\alpha_0}(\cdot,\cdot)$ and hence is identical to the channel from $\mathcal{F}^1_{N(\alpha_0)}$ to $\mathcal{F}^2_{\{\alpha_0\}}$. If we then consider the definition (2.6) of $\rho(\alpha_0)$ with the roles of m_0 and m_1 in that definition being played by

$$\mathbf{P}[X_n \in \cdot \mid X_{n+1,\alpha} = x_\alpha;\ \alpha \in B_1,\ X_{n,\beta} = y_\beta, \beta \in B_2, X_0 = j]\,,$$

$j = 0, 1$, then (3.3) follows and the proof is complete. □

Lemma 3.2.

$$I(\mathcal{F}_0 \wedge \mathcal{F}^n_{N(\alpha_r)} \mid \mathcal{F}^{n+1}_{\{\alpha_0, \ldots, \alpha_{r-1}\}} \vee \mathcal{F}^n_K) =$$
$$= I(\mathcal{F}_0 \wedge \mathcal{F}^n_{N(\alpha_r)} \mid \mathcal{F}^n_K) - I(\mathcal{F}_0 \wedge \mathcal{F}^{n+1}_{\{\alpha_0, \ldots, \alpha_{r-1}\}} \mid \mathcal{F}^n_K)$$
$$+ I(\mathcal{F}_0 \wedge \mathcal{F}^{n+1}_{\{\alpha_0, \ldots, \alpha_{r-1}\}} \mid \mathcal{F}^n_{N(\alpha_r)} \vee \mathcal{F}^n_K)\,. \tag{3.4}$$

$$I(\mathcal{F}_0 \wedge \mathcal{F}^{n+1}_{J\cup\alpha} \mid \mathcal{F}^n_K) \leqslant \rho I(\mathcal{F}_0 \wedge \mathcal{F}^{n+1}_J \mid \mathcal{F}^n_{N(\alpha)\cup K}) + \rho I(\mathcal{F}_0 \wedge \mathcal{F}^n_{N(\alpha)} \mid \mathcal{F}^n_K)$$
$$+ (1-\rho) I(\mathcal{F}_0 \wedge \mathcal{F}^{n+1}_J \mid \mathcal{F}^n_K)\,. \tag{3.5}$$

Proof. We have

$$I(\mathcal{F}_0 \wedge \mathcal{F}^n_{N(\alpha_r)} \mid \mathcal{F}^{n+1}_{\{\alpha_0, \ldots, \alpha_{r-1}\}} \vee \mathcal{F}^n_K) =$$
$$= I(\mathcal{F}_0 \wedge \mathcal{F}^n_{N(\alpha_r)} \mid \mathcal{F}^n_K) - I(\mathcal{F}^n_0 \wedge \mathcal{F}^n_{N(\alpha_r)} \wedge \mathcal{F}^{n+1}_{\{\alpha_0, \ldots, \alpha_{r-1}\}} \mid \mathcal{F}^n_K)$$
$$= I(\mathcal{F}_0 \wedge \mathcal{F}^n_{N(\alpha_r)} \mid \mathcal{F}^n_K) - I(\mathcal{F}_0 \wedge \mathcal{F}^{n+1}_{\{\alpha_0, \ldots, \alpha_{r-1}\}} \mid \mathcal{F}^n_K)$$
$$+ I(\mathcal{F}_0 \wedge \mathcal{F}^{n+1}_{\{\alpha_0, \ldots, \alpha_{r-1}\}} \mid \mathcal{F}^n_{N(\alpha_r)} \vee \mathcal{F}^n_K)$$

so that (3.4) follows.

By Lemma 3.1,

$$I(\mathcal{F}_0 \wedge \mathcal{F}^{n+1}_{\{\alpha\}} \mid \mathcal{F}^{n+1}_J \vee \mathcal{F}^n_K) \leqslant \rho I(\mathcal{F}_0 \wedge \mathcal{F}^n_{N(\alpha)} \mid \mathcal{F}^{n+1}_J \vee \mathcal{F}^n_K).$$

Therefore

$$\begin{aligned} I(\mathcal{F}_0 \wedge \mathcal{F}^{n+1}_{J\cup\alpha} \mid \mathcal{F}^n_K) &= I(\mathcal{F}_0 \wedge \mathcal{F}^{n+1}_J \mid \mathcal{F}^n_K) + I(\mathcal{F}_0 \wedge \mathcal{F}^{n+1}_{\{\alpha\}} \mid \mathcal{F}^{n+1}_J \vee \mathcal{F}^n_K) \\ &\leqslant I(\mathcal{F}_0 \wedge \mathcal{F}^{n+1}_J \mid \mathcal{F}^n_K) + \rho I(\mathcal{F}_0 \wedge \mathcal{F}^n_{N(\alpha)} \mid \mathcal{F}^{n+1}_J \vee \mathcal{F}^n_K). \end{aligned}$$

Using (3.4) we then obtain

$$\begin{aligned} I(\mathcal{F}_0 \wedge \mathcal{F}^{n+1}_{J\cup\alpha} \mid \mathcal{F}^n_K) &= I(\mathcal{F}_0 \wedge \mathcal{F}^{n+1}_J \mid \mathcal{F}^n_K) + \rho I(\mathcal{F}_0 \wedge \mathcal{F}^n_{N(\alpha)} \mid \mathcal{F}^n_K) \\ &\quad - \rho I(\mathcal{F}_0 \wedge \mathcal{F}^{n+1}_J \mid \mathcal{F}^n_K) + \rho I(\mathcal{F}_0 \wedge \mathcal{F}^{n+1}_J \mid \mathcal{F}^n_{N(\alpha)\cup K}) \\ &= (1-\rho) I(\mathcal{F}_0 \wedge \mathcal{F}^{n+1}_J \mid \mathcal{F}^n_K) + \rho I(\mathcal{F}_0 \wedge \mathcal{F}^n_{N(\alpha)} \mid \mathcal{F}^n_K) \\ &\quad + \rho I(\mathcal{F}_0 \wedge \mathcal{F}^{n+1}_J \mid \mathcal{F}^n_{N(\alpha)\cup K}) \end{aligned}$$

so that the proof is complete. □

We now introduce the shorthand notation

$$I^+_n(J \mid K) = I(\mathcal{F}_0 \wedge \mathcal{F}^n_J \mid \mathcal{F}^{n-1}_K),$$

$$I_n(J \mid K) = I(\mathcal{F}_0 \wedge \mathcal{F}^n_J \mid \mathcal{F}^n_K), \quad n = 1, 2, 3, \ldots .$$

Note that $I_n(J \mid K) \geqslant 0$ and $I^+_n(J \mid K) \geqslant 0$. We can rewrite (3.5) as

$$I^+_{n+1}(J \cup \alpha \mid K) \leqslant (1-\rho) I^+_{n+1}(J \mid K) + \rho I^+_{n+1}(J \mid K \cup N^{(\alpha)}) + \rho I_n(N(\alpha) \mid K). \tag{3.6}$$

Given a set function $E_1(\cdot)$, we set

$$E_1(J \mid K) = E_1(J \cup K) - E_1(K), \quad E^+_1(J \mid K) = E_1(J)$$

and for $n \geqslant 1$ we define $E^+_{n+1}(\cdot \mid \cdot)$ and $E_{n+1}(\cdot \mid \cdot)$ recursively as follows:

$$\begin{aligned} E^+_{n+1}(J \cup \alpha \mid K) &\equiv (1-\rho) E^+_{n+1}(J \mid K) + \rho E^+_{n+1}(J \mid K \cup N(\alpha)) + \rho E_n(N(\alpha) \mid K) \\ E_{n+1}(J \mid K) &= E^+_{n+1}(J \cup K \mid \emptyset - E^+_{n+1}(K \mid \emptyset) \\ E_n(J) &\equiv E_n(J \mid \emptyset). \end{aligned} \tag{3.7}$$

Theorem 3.3.

$$E_2^+(\alpha_1 \cup \ldots \cup \alpha_N \mid K) = \tag{3.8}$$

$$= \sum_{k=0}^{N-1} [\rho^{N-k}(1-\rho)^k \sum_k E_1(N(\alpha_{i_1}) \cup \ldots \cup N(\alpha_{i_{N-k}}) \mid K)]$$

where for $k = 1, \ldots, N-1$, $\sum_k$ is the sum over all subsets of $\{\alpha_1, \ldots, \alpha_N\}$ having $(N-k)$ elements.

Proof. By (3.7)

$$E_2^+(\alpha_1 \mid K) = \rho E_1(N(\alpha_1) \mid K)$$

$$E_2^+(\alpha_1 \cup \alpha_2 \mid K) = (1-\rho) E_2^+(\alpha_1 \mid K) + \rho E_2^+(\alpha_1 \mid K \cup N(\alpha_2)) + \rho E_1(N(\alpha_2) \mid K)$$

$$= (1-\rho)\rho E_1(N(\alpha_1) \mid K) + \rho^2 E_1(N(\alpha_1) \mid K \cup N(\alpha_2)) + \rho E_1(N(\alpha_2) \mid K)$$

$$= (1-\rho)\rho [E_1(N(\alpha_1) \mid K) + E_1(N(\alpha_2) \mid K)] + \rho^2 E_1(N(\alpha_1) \cup N(\alpha_2) \mid K).$$

We now proceed by mathematical induction. Let us assume that for all K,

$$E_2^+(\alpha_1 \cup \ldots \cup \alpha_N \mid K) = \sum_{k=0}^{N-1} \left[\rho^{N-k}(1-\rho)^k \sum_k E_1(N(\alpha_{i_1}) \cup \ldots \cup N(\alpha_{i_{N-k}}) \mid K)\right]$$

Then by (3.7)

$$E_2^+(\alpha_1 \cup \ldots \cup \alpha_{N+1} \mid K) = (1-\rho) E_2^+(\alpha_1 \cup \ldots \cup \alpha_N \mid K)$$

$$+ \rho E_2^+(\alpha_1 \cup \ldots \cup \alpha_N \mid K \cup N(\alpha_{N+1})) + \rho E_1(N(\alpha_{N+1} \mid K)$$

$$= (1-\rho)\left[\sum_{k=1}^{N-1} \rho^{N-k}(1-\rho)^k \sum_k E_1(N(\alpha_{i_1}) \cup \ldots \cup N(\alpha_{i_{N-k}}) \mid K)\right]$$

$$+ \rho\left[\sum_{k=0}^{N-1} \rho^{N-k} \sum_k E_1(N(\alpha_{i_1}) \cup \ldots \cup N(\alpha_{i_{N-k}}) \mid K \cup N(\alpha_{N+1}))\right]$$

$$+ \rho E_1(N(\alpha_{N+1}) \mid K).$$

Noting that

$$\rho\left[\sum_{r=0}^{N} \rho^{N-r}(1-\rho)^r \binom{N}{r}\right] = \rho,$$

and that

$$E_1(N(\alpha_{i_1}) \cup \ldots \cup N(\alpha_{i_{N-k}}) \mid K \cup N(\alpha_{N+1})) + E_1(N(\alpha_{N+1}) \mid K) =$$

$$= E_1(N(\alpha_{i_1}) \cup \ldots \cup N(\alpha_{i_{N-k}}) \cup N(\alpha_{N+1}) \mid K)$$

we obtain

$$E_2^+(\alpha_1 \cup \ldots \cup \alpha_{N+1} \mid K) =$$

$$= (1-\rho)\left[\sum_{k=0}^{N-1} \rho^{N-k}(1-\rho)^k \sum_k E_1(N(\alpha_{i_1}) \cup \ldots \cup N(\alpha_{i_{N-k}}) \mid K)\right]$$

$$+ \rho(1-\rho)^N E_1(N(\alpha_{N+1}) \mid K)$$

$$+ \rho\left[\sum_{k=0}^{N-1} \rho^{N-k}(1-\rho)^k \sum_k E_1(N(\alpha_{i_1}) \cup \ldots \cup N(\alpha_{i_{N-k}}) \cup N(\alpha_{i_{N+1}}) \mid K)\right].$$

Hence combining terms we obtain

$$E_2^+(\alpha_1 \cup \ldots \cup \alpha_{N+1} \mid K) = \sum_{k=0}^{N}\left[\rho^{N+1-k}(1-\rho)^k \sum_k E_1(N(\alpha_{i_1}) \cup \ldots \cup N(\alpha_{i_{N+1-k}}) \mid K)\right]$$

and the proof is complete. □

Corollary 3.4. *If $E_1^1(J) \geqslant E_1^2(J)$ for all J, then $E_n^1(J) \geqslant E_n^2(J)$ for all J and all n.*

Proof. Note that for $i = 1, 2$,

$$E_{n+1}^i(\alpha_1 \cup \ldots \cup \alpha_N) = \sum_{k=0}^{N-1}\left[\rho^{N-k}(1-\rho)^k \sum_k E_n^i(N(\alpha_{i_1}) \cup \ldots \cup N(\alpha_{i_{N-k}}))\right].$$

Since all coefficients are non-negative the result is obvious. □

Corollary 3.5. *If $E_1(\cdot \mid \emptyset) = I_1(\cdot \mid \emptyset)$, then $I_n(J \mid \emptyset) \leqslant E_n(J \mid \emptyset)$ for all n and J.*

Proof. Retracing the steps of the proof of Theorem 3.3 and using inequality (3.6) we can show

$$I_2^+(\alpha_1 \cup \ldots \cup \alpha_N \mid K) \leqslant \sum_{k=0}^{N-1}\left[\rho^{N-k}(1-\rho)^k \sum_k I_1(N(\alpha_{i_1}) \cup \ldots \cup N(\alpha_{i_{N-k}}) \mid K)\right].$$

Hence

$$I_2(\alpha_1 \cup \ldots \cup \alpha_N) \leqslant \sum_{k=0}^{N-1} \left[\rho^{N-k}(1-\rho)^k \sum_k I_1(N(\alpha_{i_1}) \cup \ldots \cup N(\alpha_{i_{N-k}})) \right].$$

Hence $I_2(J) \leqslant E_2(J)$. Let Ψ be the transformation from E_n to E_{n+1} and proceed by induction. If $I_n(J) \leqslant E_n(J)$ for all J, then

$$I_{n+1}(\cdot) \leqslant \Psi(I_n(\cdot)) \leqslant \Psi(E_n(\cdot)) = E_{n+1}(\cdot)$$

and the proof is complete. □

Theorem 3.6. *If* $|N(\alpha)| \leqslant m$ *for each* α *and* $m\rho < 1$, *then* $I_n(J) \to 0$ *as* $n \to \infty$ *for each* $J \subset A$.

Proof. Without loss of generality we can assume that

$$E_1(\alpha_1 \cup \ldots \cup \alpha_k) \leqslant k\gamma$$

where γ is a constant. Then

$$E_2(\alpha_1 \cup \ldots \cup \alpha_N) \leqslant \sum_{k=0}^{N-1} \left[\rho^{N-k}(1-\rho)^k \sum_k E_1(N(\alpha_{i_1}) \cup \ldots \cup N(\alpha_{i_{N-k}})) \right]$$

$$\leqslant \sum_{k=0}^{N-1} [\rho^{N-k}(1-\rho)^k \tbinom{N}{k} (N-k)\gamma m]$$

$$= N\gamma m\rho \sum_{k=0}^{N-1} \rho^{(N-1)-k}(1-\rho)^k \tbinom{N-1}{k}$$

$$= N\gamma\rho m(\rho + (1-\rho))^{N-1} = \gamma N(\rho m).$$

Hence

$$E_n(\alpha_1 \cup \ldots \cup \alpha_N) \leqslant \gamma N(\rho m)^{n-1}.$$

Hence if $\rho m < 1$, $E_n(\alpha_1 \cup \ldots \cup \alpha_N) \to 0$ as $n \to \infty$ and the proof is complete. □

Corollary 3.7. *Under the assumptions of Theorem* 3.6, *the synchronous Markov system is ergodic.*

Proof. Refer to [4]. □

Let us return to the equation

$$E_{n+1}(\alpha_1 \cup \ldots \cup \alpha_N) = \sum_{k=0}^{N-1} \sum_k E_n(N(\alpha_{i_1}) \cup \ldots \cup N(\alpha_{i_{N-k}}))] \, \rho^{N-k} (1-\rho)^k. \tag{3.9}$$

We will now find a probabilistic interpretation. Let $\bar{\mathcal{P}}(A)$ denote the class of finite subsets of A. Consider a Markov chain $[Y_n]$ with state space $\bar{\mathcal{P}}(A)$ and with transition probabilities given by

$$\begin{aligned} P_{JL} &= \sum_{\substack{K \subset J \\ K \cup \partial K = L}} \rho^{|K|}(1-\rho)^{|J|-|K|} && \text{if } L \cap (J \cup \partial J)^c = \emptyset \\ &= 0 && \text{if } L \cap (J \cup \partial J)^c \neq \emptyset \end{aligned}$$

where $\partial K \equiv [\alpha \colon \alpha \in N(\beta), \beta \in K]$. In other words a random subset of J is chosen and then all the boundary points are added. (3.9) can then be rewritten as

$$E_{n+1}(J) = \sum_L P_{JL} \, E_n(L) \equiv (TE_n)(J) \, .$$

Hence

$$E_{n+1}(J) = (T^n E_1)(J) \, . \tag{3.10}$$

Note that $\emptyset$ is an absorbing state for the Markov chain. Let

$$\rho^* \equiv \sup\{\rho \colon \mathbf{P}_\rho[Y_n \to \emptyset \text{ as } n \to \infty] = 1\} \, . \tag{3.11}$$

Note that if $\rho < \rho^*$, that is, the system is "subcritical", then $I_n(J) \to 0$ as $n \to \infty$ for all J and the system is ergodic.

Theorem 3.8 (Bucy [1], Holley [7]). *Assume that there exists a function $f(\cdot)$ on $\bar{\mathcal{P}}(A)$ satisfying*

$$f(\emptyset) = 0 \, , \tag{3.12}$$

$$f(\cdot) \geqslant 0 \, , \tag{3.13}$$

$$f(B) \to \infty \text{ as } |B| \to \infty \, , \tag{3.14}$$

$$f(\cdot) \text{ is superharmonic, that is, } f(B) \geqslant \Sigma P_{BL} \, f(L) \, . \tag{3.15}$$

Then $\mathbf{P}[Y_n \to \emptyset \text{ as } n \to \infty] = 1$.

Proof. Note that $\mathbf{P}[Y_n \to \emptyset \text{ as } n \to \infty] = 1$ if and only if $\mathbf{P}[\limsup_{n\to\infty} |Y_n| = \infty] = 0$. If $f(\cdot)$ is superharmonic, then $f(Y_n)$ is a non-negative supermartingale. But every non-negative supermatingale converges almost surely to a finite valued function. But then

$$\mathbf{P}\left[\limsup_{n\to\infty} |Y_n| = \infty\right] = 0$$

and the proof is complete. □

The following examples are closely related to those presented by Holley in [7].

Example 3.9. Let $A = \mathcal{T}^3$ (the 3-tree). Given a finite set B, a point $\alpha \in B$ is said to be an end point if we can travel outwards from B in two directions without returning to B. We then define

$$f(B) = \text{Number of end points in } B\,.$$

Given Y_n we must compute the value of $f(Y_{n+1})$. When an endpoint is included in the random subset, this increases by 2 the number of such endpoints. When an endpoint is excluded then there is a decreases of 1 if the neighbouring interior point is also excluded, otherwise the net change is zero. Hence

$$\sum_L P_{JL}\, f(L) \leqslant f(J)\,[1 + 2\rho - (1-\rho)^2]\,.$$

But then $f(\cdot)$ is superharmonic if $1 + 2\rho - (1-\rho)^2 \leqslant 1$, that is, if $\rho \leqslant 2 - \sqrt{3} \doteq .2679$. Hence $\rho^* \geqslant .2679$. Note that this is an improvement on the estimate $\rho^* \geqslant .25$ obtained from Theorem 3.6.

Example 3.10. Let $A = \mathbf{Z}$ and $N(\alpha) \equiv \{\alpha - 1, \alpha, \alpha + 1\}$. In this case let $f(B) \equiv$ diameter of B. Again we consider the computation of $f(Y_{n+1})$ given Y_n. If an end point is included in the random subset then the diameter increases by 1. If an end point is excluded then the length decreases by 1 or more if its interior neighbour is also excluded. Hence if $f(J) \geqslant 4$, then

$$\sum_L P_{JL}\, f(L) \leqslant f(J) + 2\rho - 2(1-\rho)^2\,.$$

Hence $f(\cdot)$ is superharmonic if $\rho - (1-\rho)^2 \leqslant 0$, that is, $\rho \leqslant (3 - \sqrt{5})/2 \doteq .3820$. Hence $\rho^* \geqslant .3820$. In [4], this result is improved by other methods and it is shown that $\rho^* \geqslant .5$.

4. Information flow in asynchronous systems

Let us now consider an asynchronous Markov system associated with the family $\{q_\alpha(\cdot,\cdot)\colon \alpha \in A\}$ and the speed measure $Q(\cdot)$. Let Ξ_n denote the random subset of A at which transitions are allowed between time n and time $n+1$. Let $\mathcal{H}_n \equiv \sigma(\Xi_n)$. Also let

$$\Psi(J) \equiv \sum_{D\subset J} |D|\,[Q(U(D) \cap L(J-D))]\,.$$

Theorem 4.1.

$$I_{n+1}(J) \leqslant \sum_{D\subset J} [Q(U(D)\cap L(J-D))] \times \tag{4.1}$$

$$\times \left\{ \sum_{k=0}^{|D|} \rho^{|D|-k}(1-\rho)^k \sum_k I_1((J-D)\cup N(\alpha_{i_1})\cup \ldots \cup N(\alpha_{i_{|D|-k}}))\right\}$$

where Σ_k represents the sum over all subsets of D containing $(|D|-k)$ elements.

Proof. $I(\mathcal{F}_0 \wedge \mathcal{F}_J^{n+1} \wedge \mathcal{H}_n) + I(\mathcal{F}_0 \wedge \mathcal{F}_J^{n+1} \mid \mathcal{H}_n) = I(\mathcal{F}_0 \wedge \mathcal{F}_J^{n+1})$. Noting that

$$I(\mathcal{F}_0 \wedge \mathcal{F}^n_{J\cup\partial J} \wedge \mathcal{H}_n) = 0, \qquad I(\mathcal{F}_0 \wedge \mathcal{F}_J^{n+1} \wedge \mathcal{H}_n \mid \mathcal{F}^n_{J\cup\partial J}) = 0\,,$$

we deduce that

$$\begin{aligned} I(\mathcal{F}_0 \wedge \mathcal{F}_J^{n+1} \wedge \mathcal{H}_n) &= I(\mathcal{F}_0 \wedge \mathcal{F}_J^{n+1} \wedge \mathcal{F}^n_{J\cup\partial J} \wedge \mathcal{H}_n) \\ &= -I(\mathcal{F}_0 \wedge \mathcal{F}^n_{J\cup\partial J} \wedge \mathcal{H}_n \mid \mathcal{F}_J^{n+1}) \\ &= -I(\mathcal{F}_0 \wedge \mathcal{H}_n \mid \mathcal{F}_J^{n+1}) + I(\mathcal{F}_0 \wedge \mathcal{H}_n \mid \mathcal{F}_J^{n+1} \vee \mathcal{F}^n_{J\cup\partial J}) \\ &= -I(\mathcal{F}_0 \wedge \mathcal{H}_n \mid \mathcal{F}_J^{n+1}) \leqslant 0\,. \end{aligned}$$

Hence

$$I(\mathcal{F}_0 \wedge \mathcal{F}_J^{n+1}) \leqslant I(\mathcal{F}_0 \wedge \mathcal{F}_J^{n+1} \mid \mathcal{H}_n)\,.$$

Hence

$$I_2(J) \leqslant \sum_{D\subset J} [Q(U(D)\cap L(J-D))]\; I_2(J \mid U(D)\cap L(J-D))\,.$$

But according to Thecrem 3.3,

$$I_2(J \mid U(D) \cap L(J-D)) \leqslant$$

$$\leqslant I_1(J-D) + \sum_{k=0}^{|D|-1} \rho^{|D|-k}(1-\rho)^k \cdot \sum_k I_1(N(\alpha_{i_1}) \cup \ldots \cup N(\alpha_{i_{|D|-k}} \mid J-D)) .$$

We then obtain the result by decomposing $I_1(J-D)$ and collecting terms.□

Theorem 4.2. *Assume that $|N(\alpha)| \leqslant m$ for each α, $m\rho < 1$ and $\Psi(J)/|J| \geqslant \eta > 0$ for all J. Then $I_n(J) \to 0$ as $n \to \infty$ for each $J \subset A$.*

Proof. Without loss of generality we can assume that $I_1(J) \leqslant \gamma |J|$ where γ is a constant. Then by Theorem 4.1,

$$I_2(J) \leqslant \sum_{D \subset J} [Q(U(D) \cap L(J-D))] \times$$

$$\times \sum_{k=0}^{|D|} \rho^{|D|-k}(1-\rho)^k \sum_k (|J-D|\gamma + (|D|-k)m\gamma)$$

$$= \sum_{D \subset J} [Q(U(D) \cap L(J-D))] \times$$

$$\times \sum_{k=0}^{|D|} \rho^{|D|-k}(1-\rho)^k (|J-D|\gamma + (|D|-k)\binom{|D|}{k} m\gamma)$$

$$= \sum_{D \subset J} [Q(U(D) \cap L(J-D))]\{|J-D|\gamma + |D|\rho m\gamma\}$$

$$= \gamma \{|J| + \sum_{D \subset J} [Q(U(D) \cap L(J-D))]\{|D|(m\rho - 1)\}$$

$$= \gamma |J| (1 + (\Psi(J)/|J|)(m\rho - 1))$$

$$\leqslant \gamma |J|(1 - \eta(1-m\rho)) = \gamma |J| \delta, \quad \delta < 1 .$$

Hence

$$I_n(J) \leqslant \delta^{n-1} \gamma |J| \text{ for all } J \text{ and } n .$$

Hence

$$\lim_{n\to\infty} I_n(J) = 0 \quad \text{for all } J ,$$

and the proof is complete. □

References

[1] R.S. Bucy, Stability and positive supermartingales, J. Differential Equations 1 (1965) 151–155.
[2] J. Chover, Convergence of a local lattice process, Univ. of Wisconsin preprint (1974).
[3] D.A. Dawson, Information flow in discrete Markov systems, J. Appl. Probability 10 (1973) 63–83.
[4] D.A. Dawson, Information flow in one dimensional Markov systems, to appear in Proc. Am. Math. Soc.
[5] D.A. Dawson, Information flow in some classes of Markov systems, to appear in J. Appl. Probability.
[6] D.A. Dawson, Synchronous and asynchronous reversible Markov systems, Carleton Math. Series No. 91 (1973).
[7] R. Holley, Lecture given at Dartmouth College, July, 1973.
[8] L.N. Vasserstein, Markovian processes on a countable product of spaces describing a large system of automata, Problemy Peredači Informacii 5 (1969) 64–72.